DOCUMENT RESUME

ED 094 007 TM 003 926

AUTHOR	Foskett, William; Young, Joseph
TITLE	Search For Success. Toward Policy on Educational Evaluation.
INSTITUTION	National Advisory Council on Education Professions Development, Washington, D.C.
PUB DATE	[Jun 74]
NOTE	41p.; For a related document, see ED 083 188
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	Decision Making; Educational Assessment; *Educational Research; *Evaluation; *Evaluation Needs; Federal Government; *Government Role; Policy; Program Evaluation

ABSTRACT

The improvement of the validity and usefulness of evaluation research will require improving both the theory and technology of evaluation, and bettering the ways in which the federal government plans, organizes, and uses evaluation research. Three general kinds of evaluation research are currently practiced: analytic, process, and product. These are appropriate to each of three phases of government: planning/budgeting, operations, and effects. The conditions necessary to improve the quality of evidence which evaluation research provides for federal decision processes rest in the overlapping realms of legislation administration and technology. As the significance of evaluation research increases, its potential to harm will also increase. Therefore certain safeguards become necessary. These safeguards are discussed. The combined efforts of the Executive and the Congress should create and sustain an educational evaluation research system yielding evidence which searches out successes, permits successes to be weighed against failures, reviews the nature and effects of successes, provides sufficient evidence for responsible decision, and asks why successes or failures occurred. (Author/SC)

ED 094007

TM 003 926

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

TM

SEARCH FOR SUCCESS

Toward Policy on Educational Evaluation

A Report of the
National Advisory Council
on Education Professions Development

June, 1974

National Advisory Council on Education Professions Development
Room 308, 1111 - 20th Street, N.W., Washington, D.C. 20036

The National Advisory Council on Education Professions Development was established by Public Law 90-35 in June of 1967. Members are appointed by the President. The Council is charged with the review of the Education Professions Development Act and of all other Federal programs for the training and development of educational personnel. Reports of findings and recommendations are made to the President and to the Congress.

Members of the Council* submitting this report are:

Lyle E. Anderson, Jr.
Chairman
School Board of Broward County
Ft. Lauderdale, Florida

Waldo R. Banks, Sr.
President, American Education
Economic Assistance Foundation
Carson, California

Larry J. Blake (Chairman)
President, Flathead Valley
Community College
Kalispell, Montana

Jason E. Boynton
Associate Professor of Education
University of New Hampshire
Durham, New Hampshire

R. Creighton Buck
Professor of Mathematics and
Acting Director, Mathematics
Research Center
University of Wisconsin-Madison
Madison, Wisconsin

Jennie A. Caruso
Dean of Women, Maple Heights West
Junior High School
Maple Heights, Ohio

Helen G. Edmonds
Dean of the Graduate School
North Carolina Central University
Durham, North Carolina

Janet C. Erickson
Community Service Volunteer
San Marino, California

Thomas R. Hills, Chairman
Junior College Division
Black Hills State College
Spearfish, South Dakota

M. Elizabeth Jacka
Executive Vice President, National
Merit Scholarship Corporation
Evanston, Illinois

Julia Mills Jacobsen
Coordinator of Government Relations
Lynchburg, Randolph-Macon Woman's,
and Sweet Briar Colleges
Washington, D. C.

Marvin D. Johnson
Vice President for University Relations
University of Arizona
Tucson, Arizona

Arnulfo L. Oliveira
President
Texas Southmost College
Brownsville, Texas

Joseph S. Radom
Jaffe, Snider, Raitt, Garrat and Heuer
Detroit, Michigan

Martin W. Schoppmeyer
Professor of Education
University of Arkansas
Fayetteville, Arkansas

---

*as of April 26, 1974

* * * * * * * * * * * * *

Joseph Young
Executive Director

Contents

PREFACE

QUICKENING CURRENTS AND CROSSCURRENTS.....1

Need for Evaluation Policy.....2

SCOPE OF EVALUATION.....4

The Setting of Evaluation.....4
Evaluation Budgeting and Priorities.....6
Differences of Authority and Interests.....7
Information and Power.....8
Capacity Building.....9
Program Policy.....10
Objectives and Evaluation.....11

CONDITIONS FOR IMPROVEMENT.....12

Legislative Conditions.....14
Legislative language.....14
Appropriation of funds.....15
Oversight of evaluation.....15
Resources.....15
Oversight of programs.....16

Administrative Conditions for Improvement.....16
Purposes and coordination in program evaluation policy.....16
Sharing the benefiet and burden of evaluation.....18
Information overload.....18
Dominance and diversity in evaluation research....19
Research quality and "outside" contracting.....20

Technological Conditions.....21
Linking services, costs and results.....21
Matching measures and statistics with services and objectives.....23
And also, the precision of poetry.....25
Capacity to search for success.....25
Research and development in evaluation.....26

Safeguards.....26
Evaluation failure or program failure?.....26
Certification of readiness.....27
Quality control in data collection and analysis...27
Correctives for bias.....28
Summary gaps.....29
Protection for evaluators, administrators and participants.....29
Meta-evaluation.....30

What Should an Educational Evaluation Research System Do?.....30
Search for success.....31
Weigh success against failure.....31
Review the nature of success.....31
Sufficiency of evidence.....32
Ask "Why?".....32

PROBLEMS

Superficial Compliance.....33
Conflicting Assumptions in Science and Civil Liberties.....33

RECOMMENDATIONS

PREFACE

Increasingly questions are being raised about a wide range of issues concerning Federal roles in education, major pieces of educational legislation and particular programs created by this legislation. Decisions on these matters can be made in two ways:

> --*on the basis of political power, partisanship, or the prestige of advocates of a particular position, or*
>
> --*on the basis of independent, objective search for evidence that establishes the merits of particular positions.*

Providing independent, objective evidence is the burden of the evaluation research activities carried out by the Federal Government. Integrity and intelligence in the use of this evidence is as essential as the validity of the evidence provided. The importance of the decisions concerning public policy questions in education defines the importance of evaluation.

Evaluation is essential to educational policy as indicated in this Council's report specifying essential elements of educational policy, Windows to the Bureaucracy. Hence, evaluation itself must be thought through and governed by general policy and by *policies of effective means* formulated to ensure that larger goals are achieved with the highest degree of effectiveness and the greatest efficiency in use of resources. Thus policies, systems and uses of evaluation must be continually examined.

The authors of this report -- Council staff members William Foskett and Joseph Young -- wish to thank the many evaluation researchers and program administrators whose comments and criticism have assisted the authors in their inquiries over the last several years. The Council and the authors are especially grateful for the several hundred thoughtful responses they received to their preliminary report, Evaluation of Education: In Need of Examination, from educators and evaluators all over the country, as well as from members of Congress.

June 1, 1974
Washington, D. C.

QUICKENING CURRENTS AND CROSSCURRENTS

The call to examine the processes and products of Federal educational policies in light of our nation's needs and purposes has echoed in every branch, level and phase of our government. As the complexities of our nation race toward the limits of our resources of understanding, time and money, and as the boundaries of our actions further overlap each other, extension of the use of evaluation research is desirable and necessary.

Evaluation research is emerging as an important tool of government. With this tool, evidence regarding the prospects or the accomplishments of Federal programs is developed using the concepts and methods of the social sciences. This evidence is mustered in the examination of program policies and public spending. Such evidence has become an important adjunct to the planning and management decisions of the Executive branch; and it is an essential ingredient for Congressional oversight and budget review processes. Evidence from educational evaluation research has entered into judicial proceedings with increasing frequency.

The evidence developed with the concepts and technology of evaluation research has carried with it an aura of authority and credibility associated with its origins in scientific (social science) research. Program evaluation strikes a posture of independence and objectivity, valued so highly that its authority vies with that of legitimate political processes for making decisions about Federal programs.

Despite the increasing expenditures, need, and significance for mustering the kind of evidence sought by evaluation research, both the quality and the use of such evidence has been quite disappointing. Evaluation of Federal programs in education has proceeded on an ad hoc, improvisational basis with little provision for the conditions necessary to ensure adequate quality or valid use of the evidence produced by evaluation research.

In its preliminary report on evaluation, this Council expressed deep concern about whether the current ideas and methods of program evaluation can bear the burden they have been expected to assume in the past and whether they can meet the challenges of our future needs. In our further inquiry into Federal educational evaluation research, we have come to doubt that our government's administrative capacity has been sufficient for proper, valid mobilization and application of program evaluation theory and technology. Thus the improvement of the validity and usefulness of evaluation research will require improving both the theory and technology of evaluation, and bettering the ways in which our government plans, organizes, and uses evaluation research. Evaluation research is a necessary tool of government; but both its quality and our government's skill and wisdom in using it within the framework of our political system must be greatly improved.

Need for Evaluation Policy. The heretofore spontaneous development and application of program evaluation should now be organized much more deliberately within the framework of our political institutions, lest this valuable tool be misused or neglected. Timely, adequate improvement in the quality and use of evaluation research will require

that evaluation in the Executive Branch and the Congress be guided by formal policy on evaluation.

Two levels of policy statements on program evaluation are appropriate:

--The first level of policy should be a general, overarching statement establishing evaluation research as a tool of government. Such a policy should thoughtfully and deliberately place evaluation in our political system by an explicit, formal policy statement. It should locate evaluation properly within the organization and procedures of the Congress and the Executive. Such a policy should set forth the larger purposes of evaluation as a tool of government and should establish its metes and bounds.

--The second level of policy on evaluation should be a *policy of effective means,** stating, establishing and providing the practical conditions necessary to realize the intent of the broader policy cited above. This policy should be reviewed by appropriate committees of the Congress before its adoption, at regular intervals thereafter, and should be modified as needed.

In the next section, "The Scope of Evaluation," we shall raise some of the substantive issues which should be dealt with in a general policy on program evaluation.

---

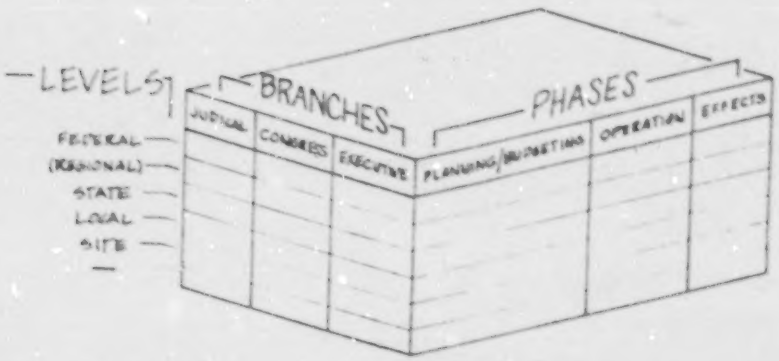
*as described in Windows to the Bureaucracy, p. 3

In the following section, "Conditions for Improvement," we shall discuss elements which we deem necessary in an evaluation policy of effective means.

SCOPE OF EVALUATION

The setting of evaluation. Evaluation occurs within a system comprised of the various branches, levels, and phases of our government, as represented in Figure 1. Evidence from evaluation

Figure 1. A Diagram of Evaluation in the Branches, Levels, and Phases of Government



studies of Federal efforts in education has emerged in all three branches of government--the Executive, the Congress and the Judiciary. At every level, the phases of governing--planning/budgeting, operations,and effects--are being examined and shaped with the techniques and evidence of evaluation research.

Three general kinds of evaluation research are currently practiced:

--analytic evaluation (or policy analysis), which includes syntheses of evaluation and research findings, pilot testing and development of policy statements;

--process (or formative) evaluation, which examines and reports on the operation or management of program activities;

--product (or impact) evaluation, which examines and reports on the quantity and quality of results obtained in educational activities.

The evidence developed by each of these kinds of evaluation is particularly appropriate to one of the three phases of government as illustrated in Figure 2 below.

Figure 2. Three General Types of Evaluation Research Most Appropriate to Each of Three Phases of Government

Phase of Government	Most Appropriate Type Evaluation Research
Planning/Budgeting	Analytic Evaluation (Policy Analyses)
Operations	Process Evaluation (Formative Evaluation)
Effects*	Product Evaluation (Impact Evaluation)

*Including costs and cost/effectiveness

When gaps or imbalances appear in some part of the system, the whole system is affected, especially the Congressional and the Executive sectors. While the functions of evaluation under the Congress and the Executive branches can be addressed by the formulation of policy, the use of evidence from evaluative research in the judiciary is a special and complex matter of jurisprudence. This Council will not comment on the use of such evidence in the Judiciary except to note that this practice is increasing, and it is of utmost importance that the sectors of government generating such evidence take every caution to ensure its quality.

Because of the interrelatedness of the environment in which evaluation occurs, merely to improve one aspect of evaluation will not be sufficient: a commitment of resources and attention to evaluation must proceed with an understanding of the whole environment of evaluation.

Evaluation, budgeting and priorities. The national budget is the most important, practical expression of national priorities and purposes. The integrity of the national budget and the debate surrounding it depend on a series of linkages between

--authorizations and appropriations

--appropriations and expenditures

--expenditures and implementation

--implementation and effectiveness.

When these linkages distort or break, the integrity of the budget is inestimably compromised and the meaningful expression of purposes and priorities is cast in doubt.

Evidence produced by evaluation research should enter the budgeting process especially to establish the linkages between expenditures and implementation, and between implementation and effectiveness. The quality and valid use of evidence from evaluation research is crucial, whether the setting is in the internal reviews of the Office of Management and Budget or in the hearings of Congressional Committees.

Differences of authority and interests. The authority of social science research is not a substitute for the authority of legitimate political institutions and interests. Despite the posture of independence, objectivity and expertise which evaluation research strikes, the evidence it yields is subject to a variety of biases inherent in the collection and statistical treatment of data, as well as to bias in the very conceptualization of the research. The evidence so developed must not be treated as if it were final and compelling truth: instead it should be treated as a legitimate viewpoint of a particular interest arrived at by following canons of scientific research. Within the canons of scientific inquiry, different and/or conflicting evidence may be produced by those with a different viewpoint.

It is a fundamental principle of our political system that conflicts of legitimate interests be resolved through adversary proceedings. Justice and fairness in decisions regarding the distribution of public resources and the nature of public institutions and programs cannot be obtained if the evidence of only one point of view or one interest is brought to bear on such decisions. Openness of such adversary proceedings to public scrutiny is necessary to inspire public confidence and credibility, as well as to encourage adequate quality in the evidence.

Information and power. The capacity to muster evidence is an essential ingredient of power in an open, rational system. It affects not only the balance of power between the three branches of our government, but the balance of power between the electorate and their government. An informed electorate is an essential ingredient of democratic government.

Growth of information-power in the Executive Branch has far exceeded that in the Congress, thereby disturbing the balance of power between the Executive and the Congress. The Executive branch has had to increase its capacity to generate, analyze and absorb evidence as the scale and complexity of our nation has grown. We believe Executive Branch evaluation capacities are not yet sufficient to meet its responsibilities in this regard. But the balance of power between the Congress and the Executive branch must be restored not by weakening the capacity of the Executive to develop, analyze and absorb information, but by increasing the capacity of the Congress to

develop, analyze, and absorb the information it needs to carry out its functions. Though Congress has begun to build its own agencies to produce evaluation research*, Congressional capacity to absorb and use the resulting evidence is still limited by current patterns of Congressional staffing, organization and procedure.

The Congressional oversight of educational programs has been brief, occasional and often narrowly focused on the fiscal propriety of program management. We reckon that much more money is lost to ineffectiveness than to avarice, and thus we feel that Congressional concern with fiscal honesty is necessary but not sufficient. Program implementation does not necessarily follow from merely legal expenditure of funds; nor does effectiveness necessarily follow from program implementation.

Resolution of Congressional/Executive information gaps will not be reached through piecemeal prescriptions of tests or research design written into legislation. It must be sought in carefully thought-through reform, examination and invention of institutions and their roles, and in regular systematic Congressional oversight of evaluation systems, as well as in appropriate legislative language.

---

*We recognize that the legislation creating these evaluation resources -- PL 91-510 and PL 92-484 -- separate responsibility for "policy analysis," "evaluation," and "technological assessment." However, our notion of evaluation includes all three inasmuch as they all have analytic, evaluative and empirical aspects addressed to providing evidence for decision-making. Inasmuch as we feel these similarities are more significant than the differences, we subsume all three under the general rubric of evaluation.

Capacity building. Much more has been called for in evaluation than we have the capacity to achieve at present; evaluation research has been oversold as well. In the short, recent history of educational evaluation research there have been considerable problems of its implementation and use. Plausible, promising concepts such as cost-effectiveness evaluation have collapsed in operation because the underlying ideas (e.g., establishing cost, measuring effectiveness) had not been developed to a practical, reliable level. Implementation of any strategy of evaluation research is also contingent on adequate supplies of personnel proficient in that strategy, on institutional accommodation, on sound fiscal policy and on other factors. There is a great gap between the idea and the reality of evaluation research. We have only begun to recognize that evaluation research is needed and must recognize that more has been expected of it than can be delivered given the limited amounts of time, attention and money thus far made available. Even with these limits, however, the sheer quantity and diversity of the evidence currently yielded by calls for evaluation would severely tax the ability of almost any decision-maker to consider and comprehend.

Program policy. In the absence of program policy which clearly states the ends and means intended in Executive branch formulations of programs it will be extremely difficult to obtain adequate evidence through evaluation research. Almost all evaluation research assumes that the objectives and logic of program plans have been worked out in advance, but this has seldom been the case. In a

previous report, Windows to the Bureaucracy, this Council has already given its views on the need for and the nature of program policy. However, such statements of program policy will not be forthcoming from the Executive branch without Congress requiring that such program policy statements be reviewed by the Congress before full-scale implementation of a program is permitted.

Foundations for oversight must be laid in advance. Before approving legislation or consenting to policy Congress should require

--a careful assessment of the needs involved

--close analysis of the logic of proposed policies or programs

--empirically documented projections of the amount and quality of resources required for success of the proposed policies or programs.

Objectives and evaluation. We encourage and support efforts to clarify the means and ends sought through policies and programs. However, great caution must be taken in heeding calls to develop "objectives" in narrow, technical terms most suitable to a particular school of thought in evaluation research. To use only those evaluation research approaches which require that objectives and logic of policies or programs be fully developed in advance places very serious limits on the opportunity to gather badly needed evidence on the nature and effectiveness of emergency or exploratory policies or programs. We must not deny ourselves the evidence needed to gauge the merits or effects of

such policies or programs; nor should we dull the ability of government to respond and learn by limiting its actions to the requirements of a particular school of thought in evaluation research. Valid evaluation can be conducted in the absence of prespecified objectives. Much might be gained if we accept the clarification of means and ends as a product rather than a precondition of valid evaluation research in such cases.

When objectives can be stated we feel it is essential that the objectives themselves be evaluated lest they provide an inadequate or invalid foundation for evaluation research or an inadequate or invalid representation of the issues pertinent to decision-makers.

CONDITIONS FOR IMPROVEMENT

The conditions necessary to improve the quality of evidence which evaluation research provides for Federal decision processes rest in the overlapping realms of legislation, administration and technology. A deliberate pace and patience will be required if legislators, administration officials and evaluation researchers are to lay a sound foundation for realizing the benefits which evaluation research can bring to our government. We identify and outline some of these conditions below.

The legislative conditions necessary to improve evaluation research evidence can be created by aggressive individual

leadership by members of Congress. The administrative and the technological conditions required for better use of better evidence from evaluation research are interrelated. There is controversy about the use and quality of evidence from evaluation research. Evaluators are charged with providing evidence that is inadequate in conceptualization, quality, timing or format. Administrators are charged with ignoring evidence that is contrary to positions taken a priori, with suppresssing evidence in interest of their own self-preservation, and with using only that evidence which supports decisions made in advance of the facts. Facts may be mustered in support of both charges, but placing blame is not a practical solution to these problems.

The administrative and the conceptual/methodological conditions needed for better evidence from evaluation research should be stated by the Executive Branch in written *policies of effective means* for evaluation research which include the elements outlined in the subsequent section below. Improvements in these domains which increase the power of evaluation research will also increase the possibilities for its misapplication. *Policies of effective means* for evaluation should anticipate these --some of which are identified here--and provide for safeguards, correctives and recourse.

The conditions necessary to improve evaluation research evidence and its use will not be achieved instantaneously, nor simultaneously. Because the domains in which these conditions

must be created overlap, the full effect of achieving them may not even be noticeable until some critical number is in place in all three domains.

Legislative Conditions

Evaluation occurs in a substantive and fiscal stage set by the Congress. Congress has exercised initiative in its early calls for evaluation in legislation. Now the Congress can provide those conditions which permit these calls to be answered and heeded by strengthening itself in the following ways.

Legislative Language. The language of the purpose and reporting sections of legislation can contribute greatly to the improvement of evaluation. More specific statement of the purposes of a particular bill can provide a set of expectations which clearly define the kind of evidence by which program failure or success should be judged. However, specific statement of legislative purpose is neither sufficient nor powerful in itself. In some cases it is necessary to frame the purposes of legislation in broad language in order to achieve its passage -- such necessity should not exempt the ensuing programs from careful evaluation. Whether a bill's purposes are stated precisely or broadly the reporting sections of a bill can provide clear guidance to the purpose, type and timing of evaluation. Inasmuch as evaluation research may serve the purposes of oversight, program management or accountability, the evaluation language of a bill should be

quite specific in regard to the purpose of evaluation.

Appropriation of funds. Sound planning, management and resource development in evaluation will require a more reliable, larger and more precisely targeted fiscal base than evaluation research has had to date. Legislative calls for evaluation must include the kinds of fiscal support and staff positions from Congress which will encourage long-range planning, and institutional continuity in those agencies of the Executive or the Congress which are assigned responsibility for evaluation research.

Oversight of evaluation. Though estimates of the amount of money spent on evaluation research vary, they are all large enough to merit Congressional oversight of the conduct of evaluation research on a regular basis. Moreover the significance of extensive use of evaluation research evidence is such that it should be kept under ongoing review. Such oversight would entail the evaluation of evaluation research -- i.e., appraisal of individual evaluation research reports, as well as the appraisal of Executive Branch evaluation research policy and its implementation.

Resources. Congress has recently created places in its Congressional agencies for the resources necessary to augment its staff in using evaluation research evidence in oversight and fiscal processes. Further general provisions for evaluation are contemplated in pending legislation (S. 1541). Now these *places* for resources must find adequate levels of support and use among the members of Congress and their staff members. Levels of support

necessary to permit Congressional agencies to sustain adequate quality of work under increasing quantity and complexity of Congressional needs are crucial to valid use of educational evaluation research in legislative processes.

Oversight of programs. Congress gives much more attention to development of new legislation than to oversight of programs ensuing from existing legislation. Introduction of evidence from evaluation research in systematic, well-established oversight hearings can bring attention to successful programs and encourage early identification and treatment of program problems. Most legislators' reputations are not established on the basis of their accomplishments in oversight, however, and thus the Congress will have to make organizational rearrangements which reward oversight with appropriate attention (e.g., a formal place in Congressional committee structure) and perquisites (e.g. staff and space) if good evidence is to be well-used. Such oversight procedures should provide a public forum in which evaluation research evidence developed from different points of view can be presented and reviewed.

Administrative Conditions for Improvement

Purposes and coordination in program evaluation-policy. The evaluation-policy for a particular program should fit the general policy of that program. Since program policy is made

within program-operating units, program evaluation-policy should be made by in-house evaluation staff with enough clout to participate effectively in planning programs which can be evaluated. However, that policy should be subject to independent, outside review before enactment.

Evaluation research may be done for a variety of purposes --e.g. budgeting/planning, improving program operations, accountability to program-clients. Program evaluation-policy should show specifically what kinds of evaluation research will be done for what reasons.

Some programs (e.g., ESEA Title I) aim to deliver a variety of services assumed necessary to reach program goals. Other programs aim to test (e.g., Follow-through) or develop (e.g., the Regional Labs and Centers) different or new educational services which will be effective. Programs which aim mainly to develop or test ideas require a different approach to evaluation than programs which aim mainly to deliver services which are assumed effective. Social experimentation to test new educational products or approaches is no substitute for evaluation of ongoing policies and programs, however.

The right kind of evaluation research should be done at the right stages of program development -- e.g., analytic evaluation should be done before and during program planning; formative evaluation is very important to do during early program implementation; and impact evaluation should be done only when a program

is fully implemented.

Evaluation designs -- especially in ongoing service-delivery programs -- must be flexible enough that sensible and necessary program changes can be made without sacrificing the evaluation. Program evaluation policy should provide for sufficient resources to track unexpected program changes and unexpected results.

Sharing the benefits and burdens of evaluation. The data collection efforts required in some Federal evaluation research has imposed burdens with little immediate benefit for the state and local agencies involved. Evaluation of Federal programs must be recognized as an interloper in state and local educational agencies. In exchange for their cooperation there is a Federal obligation to recognize, cooperate with, and to some extent, to underwrite state and local evaluation needs created by Federal activity.

Information overload. The amount, range and complexity of evidence from evaluation research tend to overwhelm decision-makers faced with diverse and immediate demands for personal attention to management or legislative duties. Improved formats and better summaries for evaluation reports are clearly needed -- but it is equally clear that alone they will be insufficient. The evidence from evaluation research must be condensed; the different lines of evidence must be drawn together (synthesized) in focus on particular issues as they arise; and in doing so,

the complexities for which decision-makers are responsible must be communicated.

Written summaries and syntheses of evaluation research evidence lag behind current conditions and must necessarily balance completeness, complexity and length against the time, training and expectations of a general audience. Better written summaries and syntheses must be backed up with an articulate staff on call to identify and explain all the evaluation research evidence bearing on a particular decision. This staff must be adequate in training and number to stay abreast of the flood of evidence. But to make it work it must be kept under review.

Dominance and diversity in evaluation research. A state government recently rejected a motorcycle safety program after an economic study found that the cost of the program would be greater than the cost of lives lost in motorcycle accidents. Such speciously narrow treatment of public problems omits moral and technical considerations which are crucial to practical public policy. Domination of the planning, conduct or interpretation of evaluation research by any particular discipline -- e.g., psychology -- or method -- e.g., econometrics -- narrows the kind of evaluative evidence on hand and thereby censors information and alternative viewpoints from decision-making. Evaluation research policy should enforce multi-disciplinary staffing at all stages of evaluation research to balance and broaden the kinds of evidence which enter into decision-making.

Research quality and "outside" contracting. Most Federal-level evaluation research is conducted by corporate or academic organizations on government contract. We seem to have drifted into this practice with an unexamined faith that it would guarantee the independence and adequate quality of evaluation research. The ensuing "independence" has not infrequently led to contract researchers pursuing their own questions rather than their client's questions. The kind of competition which would maximize the quality of evaluation research has not emerged. Better service and better quality from evaluation research contractors can be obtained by remedying the following problems.

Within limited evaluation funding on the whole, funds for contract research are so overbudgeted and in-house staff positions and funds are so underbudgeted, that good, practical conceptualization and monitoring of evaluation research is thwarted. In some cases even requests-for-proposals must be developed by outside contractors before proposals for research can be called for. Seldom are sufficient time and talent devoted to developing requests-for-proposals -- in-house "r.f.p.'s" for expenditures of millions of dollars may be assembled in only several weeks. Contractors often may have in effect only two weeks to respond to extremely complex requests. Requests-for-proposals have often attempted to specify the evaluation research design rather than the questions or issues to be addressed, or the final character of the research evidence. When a contract is let, it is often

difficult to monitor carefully -- because of vagueness, unanticipated problems or insufficient staff time. The costs of evaluation research are apparently difficult for contractors to project: representatives of two major evaluation research organizations recently declined to estimate the cost to carry out the research designs they were developing on contract as a basis for "r.f.p.'s" because attaching the price-tag was not part of their contract!

Technological Conditions

Linking services, costs and results. In general, evaluation research has focused so closely on the results of programs that very little attention has been given to good description and accurate measurement of program services. The logical and empirical linkages between proposed program services (e.g., buying library books) and program objectives (e.g., increasing mathematics achievement) should be worked out, plainly stated and evaluated analytically for their validity in the early stages of program planning. The kinds of program descriptions necessary to "sell" an approach rarely describe educational services in a way which permits outsiders to accurately replicate or understand the nature of services intended. There are many different ways services are delivered under rubrics such as "team teaching" or "individualization." Thus the services a program intends to deliver should be described fully enough that they can be measured

at the point of impact on the learner. We should know, for example, how much instructional contact "instructional aides" have with learners.

The cost-categories used for fiscal accounting tell very little about the nature of program services as delivered. Though there are measures which vary with expenditures of educational systems -- e.g., pupil/teacher ratio -- they are so far removed from the actual delivery of services to learners that there is little justification to juxtapose these measures to student achievement. Thus costs assigned to educational services should derive directly from measurable descriptions of actual educational activities. "Fiscal surrogates" -- e.g., superintendent's salary, bonded indebtedness per pupil -- which have been used to estimate effective costs for educational services bear no direct relationship to those services -- especially instructional services.

The definition of success for a program should be very carefully related to the time and ways in which program participants can be expected to grow. Evaluation research thus far has not taken into account individual differences in the rates and ways learners grow -- e.g., different children learn to read at different times in different ways at different proficiency. Nor has evaluation research taken into account that particular kinds of educational services seek different kinds of success at different times.

Cognitive achievement -- e.g., math and reading -- is a necessary but not sufficient criterion of success. When success is judged along just *one* type of measure, program resources and practices tend to focus on that one domain at the expense of quality in other services which may be necessary as well. Cognitive achievement has the broadest consensus as a criterion of program success -- but the goals of schooling are much broader and more elusive to measure than cognitive achievement.

Legitimate criteria of success used in evaluation research must represent all the stated aims of a program. For example, the effectiveness of a program aimed to deliver educational and support services to migrant workers (Title IIIB, EOA) was judged unsuccessful on the criterion of local project fiscal management and subsequently some local projects and significant program services (e.g., legal services) were terminated. The quality of local fiscal management is scarcely an expected outcome of this "allied services" program in miniature -- indeed, a proper evaluation research design could have gained us much knowledge about "allied services" approaches.

Matching measures and statistics with services and objectives. Just as the broad evaluation research strategies must match the broad strategies of program plans, so must the details of particular evaluation research designs match the details of particular programs. The instruments used should be sensitive, representative and valid measures of the program services and objectives. The statistical analyses should search out patterns

and extremes of services and results -- as well as the statements of the "average" ordinarily found.

Evaluation research plans can fail in implementation just as programs can fail. Time must be taken to "walk through" evaluation research designs by applying them to simulated program data, operational settings and decision questions, before applying the evaluation research design to the actual program. Such a process could create much more understanding of various models of evaluation research and alert decision-makers to potential problems and results in advance.

The terms in which data are collected, analyzed and reported should be selected for their congruence with the services and objectives of the program -- not for the convenience of the researcher, nor to save those with a stake in the results from the pain of bad news. Data should be collected, analyzed and reported in units of analysis which jibe with accountable decision-making units -- not in contrived regional units which comprise no specific, responsible jurisdiction. Service-delivery units of analysis (e.g., sixth-grade reading instruction) should match the units of analysis for results (e.g., sixth-grade reading scores).

The assumptions implicit in the details of the evaluation research model (e.g., that a "no-treatment" group really gets no services) and its measures (e.g., "normal" distribution of scores) should be examined for their consistency with the assumptions underlying program services and objectives. Often the

assumptions of the evaluation research design details and those of the program just don't match, and thus the evaluative evidence must be qualified -- or even discarded.

And also, the precision of poetry. Decisions which are made on only the basis of dry statistics do not adequately represent that elusive, crucial dimension of the public interest called "the quality of life." We do not mean to call for syrupy anecdotes which paint an exception as the rule. Rather, we call for including illustrative vignettes with the brevity and precision of poetry to evoke a feeling for what educational services, successes and failures mean in human terms.

Capacity to search for success. Current capacity to find and capitalize on successful projects has very serious limitations of reliability and validity. The theory and technology of this aspect of evaluation are relatively untried. Though a rudimentary technology for identifying successes is at hand, reliable data pools matching definitions of success even in terms of cognitive achievement are not generally available. The use of definitions of success with more than one criterion presents conceptual and technical problems which must be solved. But the most crucial question to be addressed is that of how to develop large-scale policy from small-scale successes. Though there are several current efforts to identify and capitalize on successful projects now under way, the theory and

technology underlying these efforts has been little improved since its first application five years ago, and may not bear the burden of its sponsors' expectations.

Research and development in evaluation. There is clear need for research and development directed toward evaluation research. Inasmuch as most evaluation research designs are now custom-tailored to every case to which they are applied, an opportunity is provided to design evaluation research with an eye not only to yielding evidence regarding its specific subject, but to yield improvements and new developments in evaluation as well. At this time the field application of evaluation designs should be organized in a manner to provide opportunities and incentives for in-the-field evaluators to participate in the advancement of evaluation theory and practice.

Safeguards

As the significance of evaluation research increases its potential to harm will also increase. Thus we feel that the following safeguards should surround the conduct and use of evaluation research.

Evaluation failure or program failure? It is possible that the best-designed and executed evaluation may fail to provide the kind and quality of evidence that is expected and needed.[7] It is inviting to construe such failure as program

failure rather than evaluation failure as in the case of the Ohio/Westinghouse evaluation of Headstart. Given the current flaws and insensitivity in evaluation research, the evidence itself must be shown valid before any judgment of program failure or success is rendered.

Certification of Readiness. The question of whether the intended activities are actually being done must be answered before the question of whether such activities are effective when they are done. Compliance with the type, quality and quantity of services intended by a policy, program or project should be certified as being delivered prior to beginning evaluation of the effects of those services. This certification should be based on the findings of process evaluation, and compliance data should be compiled for the duration of the evaluation of program effectiveness. Such certification should be publicly made to the Congress and appropriate officers of the Executive within a reasonable, specified-in-advance period of time.

Quality control in data collection and analysis. The circumstances under which evaluation research must be conducted are much more challenging than academic research. Nevertheless there are unnecessary lapses of good practice in application of social science data collection and analysis techniques in evaluation research. Despite the circumstances, the importance of the decisions at hand should not permit the use of unvali-

dated measurement instruments, unrepresentative samples, dubious application of cross-sectional designs for making longitudinal inferences about dynamic processes, neglect of distribution statistics, gross misapplication of multiple regression analysis and equally gross overinterpretation of its results, reanalysis of obsolete data for extrapolation to changed situations, and use of reporting formats which invite misinterpretation.

Minimal standards and practices should be explicitly stated in every program-specific evaluation policy. Adherence to those practices should be monitored by appropriate agencies of the Executive branch and the Congress.

Correctives for bias. The findings of evaluation research must not be cloaked with assumptions of objectivity. Evaluation research may be biased in the way it is conceived, conducted and reported. It seems impossible to avoid such biases -- they may even be necessities. Evaluation policy must provide correctives for the underrepresentation of differing social, organizational and intellectual perspectives on evaluation research and the evidence it yields. There are a variety of correctives which are yet untested and relatively undeveloped -- e.g., "legal" models of evaluation research, "goal-free" evaluation, competing analyses of the same data -- which should be tried on an experimental basis as a prelude to

treatment of this problem in policy.

Summary gaps. The various kinds of summary attempt to deal with information overload -- the flood of evidence which is overwhelming in quantity and complexity. But much can be lost in translating and condensing the findings of evaluation research reports. We have found significant gaps between the body of a report and its executive summary, between the executive summary and the news story, and between the body of the news story and its headline. Thus the heavy reliance which is now placed on written summary is an unreliable solution to the problem of information overload.

Protection for evaluators, administrators and participants. Those with a stake in the conduct and findings of evaluation research may be agents or victims in the abuse of evaluation research evidence. Those touched by the information-power of such evidence are justifiably uneasy about the legitimacy of an activity governed by rules not clearly and publicly spelled out -- especially when evaluation must be designed and conducted by "experts". Evaluators often feel the pressure of clients who feel threatened by the presence or results of evaluation. Persons who are not the particular clients of specific evaluation research often feel that evaluation research which does not represent their special perspective on events is illegitimate and unfair. Public statement of evaluation policies can go far to discourage the abuse of evaluation research. Giving

those with a stake in the evidence a voice in the definition of success and the setting of standards and practices can lend legitimacy and credibility to evaluation research evidence. Development of correctives for bias can further limit the damage done as even the best tools of government occasionally run amok.

Meta-evaluation. Little attention has been accorded the quality of evaluation research, the quality of evidence it yields, and the validity of the whole system of evaluation which has evolved in the last decade. Evaluation itself must be evaluated by evaluators and by its users in the Congressional and Executive branches of government. Evaluation policies at the agency and departmental level should provide specifically for meta-evaluation (evaluation of evaluation) to foster better internal use and to permit external (e.g., GAO, OMB) review of evaluation as a management function. Independently written appraisals of the quality and validity of evaluation research reports should be a standard feature of the report and its summary; and no report should be released for use without such appraisal.

What Should An Educational Evaluation Research System Do?

The combined efforts of the Executive and the Congress should create and sustain an educational evaluation research

system yielding evidence which

--searches out successes

--permits successes to be weighed against failures

--reviews the nature and effects of successes

--provides sufficient evidence for responsible decision

--asks why successes or failures occurred

Search for success. Capitalization on success should be the keystone of education policy development. A necessary condition for this is evaluation policy displaying in concrete terms,

--the capacity to identify and sort out successes and failures

--the capacity to identify the conditions associated with success (or failure) and to translate them into operational terms.

There are, however, major conceptual and technological elements of these tasks which must be developed and put in place.

Weigh success against failure. Success is only part of the picture. The extent of success must be compared to the extent of failure and weighed in light of the adequacy of alternative courses of action. A necessary condition for this is an evaluation system

--equally capable of revealing success or failure

--capable of mustering comparable evidence on alternative courses of action.

Review the nature of success. Is success satisfactory?

Given a successful program, does its outcome satisfy our wants and needs? Does this success move us towards goals still sought, along ways we find worthwhile? Or is the "cure" worse than the problem? Such questions should be reviewed openly in the political forums of the Executive and the Congress. A necessary condition for this is an evaluation system capable of

--translating reliable findings into common language

--providing an analytic staff sufficient to keep the Executive and Congress abreast of the quantity, as well as the concepts and methodology of evaluation.

Sufficiency of evidence. Evidence brought to bear on a decision must be commensurate in validity and reliability to the significance of action to follow from that decision. A necessary condition for having sufficient evidence in such situations is an evaluation system capable of

--foreseeing significant Federal decision situation

--gathering data, analyzing data and maintaining data and analyses in an up-to-date, useful format

Ask "Why?" Reliable evidence of success or failure is essential, but without understanding of reasons for results, policy and program improvement can proceed only by trial and error, if at all. Evaluation research has usually asked only if there were success or failure. To ask why opens a whole new question to evaluation research which will require much more time and

support. However, such costs must be weighed against costs of proceeding by trial and error.

PROBLEMS

Superficial Compliance

The processes of government are remarkably obdurate. Recent efforts to introduce so-called "rational" planning, budgeting and decision systems in government have changed primarily the rhetoric of those processes. Superficial compliance is an extraordinarily effective defense that is especially oppressive to open debate and resolution of important, legitimate issues. The implicit and explicit positions we have taken in this report represent a point of view which will be best taken if it is sufficiently robust to stimulate and survive open consideration and debate before being embraced.

Conflicting Assumptions in Science and Civil Liberties

When do the assumptions underlying social scientific evaluation research and decisions following from it become antithetical to the assumptions underlying civil liberty, individual rights and personal privacy? Hard-won freedom must not be surrendered to cadres of "experts", but that freedom will surely not be preserved by ignorance. Extension and improvement of the search for success in Federal policies and programs through evaluation research should proceed without further delay, but with deliberate speed and caution.

RECOMMENDATIONS

The recent development of educational evaluation research has been shaped almost exclusively by the Executive Branch in a setting of minimal and occasional Congressional attention. Complex organizational structures and technology for educational evaluation research have evolved in the Executive Branch which now exceed the capacity of the Congress to oversee or to use well the educational evaluation research done by the Executive Branch. Through recent legislation* Congress has begun slowly building general analytic resources in Congressional agencies; however, it has not yet begun to fully use such resources in overseeing, supporting and using educational evaluation research. It is our judgment that the next major steps to improve educational evaluation research are the responsibility of the Congress.

HENCE WE RECOMMEND THAT:

1. *Congress enact legislation requiring that the Executive Branch develop written policy to govern Executive Branch activities in educational evaluation and that such policy be submitted to systematic Congressional oversight;*

2. *Congress strengthen the legislative background of educational evaluation research -- the legislative history, language, Congressional inquiries, and the authorization and appropriation of staff positions and funding for Executive evaluation activities;*

3. *Congress conduct systematic, regular oversight hearings on the manner and quality of evaluation research done by the Executive Branch;*

4. *Congress begin to develop internal Congressional awareness, procedures, and structures for using fully the resources of Congressional Research Service, the General Accounting Office, and the Office of Technology Assessment in its oversight and use of educational evaluation research.*

Some particulars of each recommendation follow on the next pages.

---

*The legislation creating these analytic resources -- PL 91-510 and PL 92-484 -- separates responsibility for "policy analysis," "evaluation," and "technology assessment." However, our notion of evaluation includes all three because their similarieies -- analytic, evaluative and empirical inquiry addressed to decision-making -- are more significant than their differences.

1. Congress should enact legislation requiring that the Executive Branch develop written policy to govern Executive Branch activities in educational evaluation; and such policy should be submitted to systematic Congressional review and oversight.

Such policy should include

- *a documented rationale for all educational evaluation activities conducted by the Executive Branch;*
- *evaluation plans set forth in terms such that the implementation and quality of evaluation can be measured along critical dimensions;*
- *administrative remedy for isolation of available evaluation staff and data resources from formulation and making of decisions;*
- *sufficient staff to carry out such policy as well as to assist others in obtaining, understanding and using evaluation evidence;*
- *treatment of the administrative and technological issues raised in this report.*

2. Congress should strengthen the legislative background of educational evaluation research -- the legislative history, language, Congressional inquiries and the authorization and appropriation of staff positions and funding for Executive educational evaluation activities.

In these areas Congressmen can:

- *develop legislative language to*
  - *clarify the general purposes and types of evaluation the Congress intends to be conducted;*
  - *provide for the independent conduct of evaluation which serves different audiences with conflicting interests;*
  - *require independent certification that critical levels of program services are provided before and during impact evaluation;*

- require evaluation of program policies and services so that failure of effort may be distinguished from failure of strategy.

- build a background of legislative history and intent in hearings and reports which

  - develop in detail the definition of success to be applied in impact evaluations of programs or policies;

  - call forth corroborative evidence from different ways of evaluating any particular program or policy to ensure that findings are not an artifact of some particular evaluation method;

  - require that the terms in which data are reported do not mask upper or lower extremes of services or effects.

3. Congress should conduct regular, systematic, oversight hearings on the manner and quality of evaluation done by the Executive Branch.

Such oversight should include

- systematic subcommittee-level hearings to assess the quality of evidence available on the implementation and effects of particular Federal educational programs and policies;

- initiation of intensive study by an agency of the Congress to identify, analyze and evaluate the Executive Branch rationale for current patterns of practice in educational evaluation;

- regular, public review of issues of evaluation policy, practice or findings bearing on Federal decisions.

4. Congress should begin to develop internal Congressional awareness, procedures and structures for using fully the resources of the Congressional Research Service, the General Accounting Office, and the Office of Technology Assessment in its oversight and use of educational evaluation research.

Such development should include

- incentives --e.g., additional staff, publicity and perquisites -- for Congressmen to use educational evaluation research systematically;

- Congressional procedures and organization which require consideration of evidence regarding the means and ends of Federal policy in substantive terms as well as in fiscal terms;

- systematic, regular updating of members of Congress and their staff on the kinds of educational evaluation research support available from the agencies of the Congress

- keeping the budgets and staff of the Congressional agencies abreast of the increasing complexity and quantity of Congressional demands on them;

- enlargement of committee staff with more analytically-trained staff members.